JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS **159**, 1–17 (1991)

# Cycles in Extensive Form Perfect Information Games

STEVE ALPERN

*London School of Economics, Department of Mathematics, Houghton Street, London WC2A 2AE, United Kingdom*

*Submitted by George Leitmann*

Received July 26, 1989

We define and analyse a new class of perfect information games. The nodes of a directed graph $G$ are partitioned into $n$ player sets. Starting at a fixed node of $G$ an infinite path is created as follows: If the current node belongs to player $k$, then player $k$ chooses any successor node. A local reward $n$-vector is assigned to every arc. The payoff corresponding to the infinite path is the long term average of the local reward vectors. Such games are called DGA games. Negative and positive results are obtained for the existence of Nash equilibria in certain types of pure strategies (e.g., stationary and automated strategies). Applications to duopoly pricing models and "surveillance games" on graphs are given. © 1991 Academic Press, Inc.

## INTRODUCTION

In the traditional theory of games the possibility of a position reoccurring is excluded by modelling extensive forms as trees—directed graphs *without* cycles. Presumably the reason for the exclusion of cycles is to ensure that infinite play is impossible and that eventually a terminal node is reached. However, there are many examples of conflict situations in which positions may reoccur. It seems advisable to enlarge the notion of an extensive form game to cover these. This paper explores some of the mathematical consequences of allowing cycles in games of perfect information. We are particularly interested in determining situations in which the Zermelo–Kuhn Theorem, on the existence of pure strategy Nash equilibria, holds for perfect information games with cycles. As we shall see, the answer depends on the nature of the pure strategies allowed. Before describing our games more formally, we briefly mention two senses in which cycles are already covered by game theory. It is certainly true that any game with cycles is equivalent to a tree whose nodes are paths of the original game. However, this tree will be infinite and no useful general results for such trees are known. Also it should be said that in the zero-sum case our games



0022-247X/91 $3.00

are special cases of "recursive" games [E] or "stochastic games" [BK], where positions can also reoccur. But those theories are not very useful for the construction of pure strategy equilibria.

The games considered in this paper will all have the following Deterministic Graph (DG) dynamics: There is a given finite directed graph $G$, together with an assignment of one of the $n$ players to each of the nonterminal nodes. The play begins at a designated starting node and the player assigned to the current node picks the next node from among its successors in the graph. For the moment, we will say that the game stops if a terminal node (i.e., one with no successors) is reached. This dynamic structure was considered by Berge [B1, B2], but with different payoffs than considered here.

Before discussing payoffs, it will be convenient to first define strategies. A pure strategy for player $k$ is any rule that assigns to any path in $G$ which ends at a player $k$ node, one of its successor nodes. A stationary strategy is a pure strategy which depends only on the final node of the path. In other words, it always chooses the same successor whenever that node is reached. An $m$-automatic strategy is one which can be played by an automaton with $m$ internal states where inputs and outputs are coded to the nodes of $G$. As the play proceeds the current node is input into the automaton and, if it belongs to the player using this automaton, the output node is played. (A 1-automatic strategy is simply a stationary strategy.)

We now discuss payoffs. Washburn [W1] considers DG games with terminal payoff, which we will call DGT games. In these games (he considers only the two person zero sum case) a payoff is assigned to every terminal node. If the play terminates at such a node, player 1 gets the payoff assigned to that node. If the play is nonterminating, the payoff is zero. The existence of stationary optimal strategies (and a value) for DGT games follows from the general theory of recursive games [R]. However, Washburn is able to construct optimal stationary strategies by a finite recursive procedure which is polynomial in the number of nodes.

In this paper we will consider a more general payoff structure where the players may have preferences over different nonterminating plays. We technically modify the dynamics so that a terminal node is replaced by a loop (arc from the node to itself). Hence all plays of the game are nonterminating, while plays which were originally (i.e., with terminal nodes) terminating are now eventually constant. A local reward vector is assigned to every arc. The $k$th coordinate of this vector is interpreted as the payment to player $k$ made whenever the play traverses this arc. The total payoff vector vector is the long term (Cesaro) average of the local reward vectors of the traversed arcs. Such games will be called Deterministic Graph Games with Time-Average Payoffs, or DGA games. Every DGT game can be made into a DGA game by assigning the terminal payoff to the associated terminal loops and assigning 0 vector payoffs to all other arcs. Plays ending at a terminal node (loop) will have an average reward tending to the vector attached to the terminal loop. Other plays will have all their average payoffs zero. Hence DGA games subsume DGT games, as well as perfect information game trees.

There are four main results on the existence of Nash equilibria in various types of pure strategies for DGA games.

1. In DGA games, a Nash equilibrium in stationary strategies may not exist. A three person non-zero sum example is given in Section 2.

2. Every DGA game has a Nash equilibrium in pure strategies. This result (Theorem 2) is proved in Sections 3–5.

3. The previous two results say that there are equilibrium strategy profiles, but they are not too simple. The third result (Theorem 3) says that there are always $m$-automated equilibrium strategies, where $m$ is the factorial of the number of nodes.

4. As noted above, zero-sum DGA games can be thought of as very special cases of stochastic games with time-average payoff. For such games the existence of stationary optimal strategies was asserted in [G], proved rigorously in [LL], and can now be seen as a special case of the conditions in [BK]. A constructive non-probabilistic algorithm for optimal strategies is given in [EM].

In addition to these general results, we give applications of DGA games to duopoly pricing (Section 7) and "recapture" and "surveillance" games on graphs (Section 8). In Section 9 we briefly discuss possible extensions to imperfect information games with cycles, and their relation to the problem of "repeated decisions" solved by Isbell [Is].

## 1. DGA Games

In this section we give a formal definition of a DGA game and of an equilibrium strategy profile. Let $G$ be a finite directed graph with arcs $A$ and nodes $N$ partitioned into "player sets" $N_k$, $k = 1, ..., n$, where $n$ is the number of players. To ensure that the play of the game can always proceed we assume that the successor map Suc: $N \to 2^n$ is never empty, where $\mathrm{Suc}(i) = \{j\colon (i, j) \in A\}$. We assume that there is a distinguished node caled 1 in $N$. A "local reward" function $\lambda\colon A \to \mathbb{R}^n$ is given, where $\lambda_k(a)$ is interpreted as the utility payoff to player $k$ made whenever arc $a$ is traversed. The data $G$, $N$, $N_k$, and $\lambda$ completely define DGA games, as described below.

A sequence of nodes $x = (x_0, x_1, ...)$ with $x_0 = 1$ and $x_{t+1} \in \mathrm{Suc}(x_t)$ will be called a "history" if infinite and a "path" if finite. The sets of all histories and paths will respectively be denoted by $H$ and $P$. If $p=(p_0,..., p_m)$ is a path, it has an ending node $e(p)=p_m$ and a length $l(p)=m$ which counts the number of its arcs. If $t$ is a non-negative integer and $x$ is a history or a path of length at least $t$ define $x^t=(x_0, x_1,..., x_t)$ to be the initial path of $x$ of length $t$. Define $P_k=\{p\in P\colon e(p)\in N_k\}$ to be the player partition of the set of paths.

A strategy $\sigma_k$ for player $k$ is a map $\sigma_k\colon P_k\to N$ satisfying $\sigma_k(p)\in \mathrm{Suc}(e(p))$. The set of strategies available to player $k$ is denoted $S_k$. A strategy profile is a $n$-tuple $\sigma=(\sigma_1, ..., \sigma_n)$ with $\sigma_k\in S_k$. A strategy profile $\sigma$ determines a unique history $h=\infty(\sigma)$ by the recursive definition $h_0=1$ and $h_{t+1}=\sigma_k(h^t)$ if $h^t\in P_k$. In order words, if the play so far is $(h_0,..., h_t)$ and the current node $h_t$ belongs to player $k$, then player $k$ chooses the next node $h_{t+1}$ among the successor nodes of $h_t$ be a method which may depend on the entire past history $(h_0, ..., h_{t-1}, h_t)$. The map $\infty\colon S_1\times\cdots\times S_n\to H$ completely describes the dynamics of a DGA game.

To describe the payoff structure of a DGA game we first extend the local reward vector $\lambda$, originally defined on arcs, to all path $p$ by the obvious definition $\lambda(p)=\sum_{t=1}^{l(p)}\lambda(p_{t-1}, p_t)$. It is also useful to have an "average reward vector" $\lambda^*$ defined on paths $p$ by $\lambda^*(p)=\lambda(p)/l(p)$. For any strategy profile $\sigma$, let $h=\infty(\sigma)$ and define

$$\bar{\Pi}(\sigma)=\overline{\lim_{t\to\infty}}\ \lambda^*(h^t),\quad \underline{\Pi}(\sigma)=\underline{\lim_{t\to\infty}}\ \lambda^*(h^t),$$

and

$$\text{(if it exits) } \Pi(\sigma)=\lim_{t\to\infty}\lambda^*(h^t)$$

The solution concept we will use for DGA games is an "equilibrium strategy profile" by which we mean a profile $\bar{\sigma}$ satisfying, for all $k$ and all $\sigma_k\in S_k$, $\bar{\Pi}_k(\bar{\sigma}_1, ..., \bar{\sigma}_{k-1}, \sigma_k, \bar{\sigma}_{k+1}, ..., \bar{\sigma}_n)\leqslant \Pi_k(\bar{\sigma})$. Note in particular that limiting average payoffs exist for all players at an equilibrium, and that it is a Nash equilibrium as long as player $k$'s utility lies between $\underline{\Pi}_k$ and $\bar{\Pi}_k$ for all $k$.

In Section 5 we will complete the description of a finite algorithm for constructing an equilibrium strategy profile for any DGA game. Part of this algorithm uses the Zermelo–Kuhn backwards recursion applied to a finite perfect information game $F$ which we define in Section 3. The other part of the algorithm involves a simple procedure for "decycling" a path, which we describe in Section 4.

## 2. Nonexistence of Stationary Equilibria

In this section we give an example of a DGA game with no equilibrium strategy profile in stationary strategies. Let $G$ be the complete directed graph on three nodes 1, 2, 3 partitioned in the natural way ($k \in N_k$) among three players. Label the six arcs as $a = (1, 2)$, $b = (1, 3)$, $r = (2, 1)$, $s = (2, 3)$, $u = (3, 2)$, $v = (3, 1)$; and the five cycles as $C_1 = ar$, $C_2 = bv$, $C_3 = us$, $C_4 = asv$, and $C_5 = bur$. Define the local reward vector as

$$\lambda(a) = (0, 0, 3) \qquad \lambda(b) = (2, 0, 0) \qquad \lambda(r) = (0, 0, 3)$$

$$\lambda(s) = 6, 0, -3) \qquad \lambda(u) = (0, 6, 0) \qquad \lambda(v) = (0, -3, 0)$$

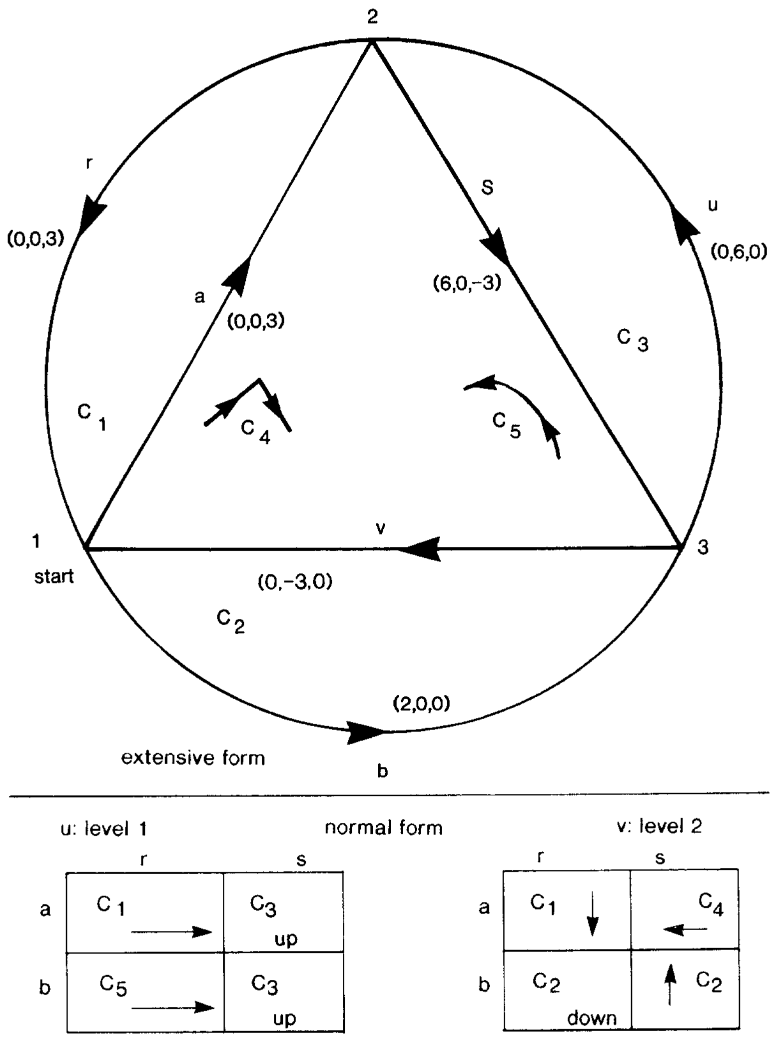


FIG. 1. A DGA game and the normal form for stationary strategies. No stationary equilibrium.

so that

$$\lambda^*(C_1) = (0, 0, 3) \qquad \lambda^*(C_2) = (1, -\tfrac{3}{2}, 0) \qquad \lambda^*(C_3)(3, 3, -\tfrac{3}{2})$$

$$\lambda^*(C_4) = (2, -1, 0)) \qquad \lambda^*(C_5) = (\tfrac{2}{3}, 2, 1).$$

All this information is depicted in Fig. 1, together with the normal form in stationary strategies (i.e., dependent on current node only, not prior history). Observe that if all players use stationary strategies then the history of the game eventually consists of repetitions of a cycle $C$ and $\pi^* = \lambda^*(C)$. Note that each of the eight boxes, corresponding to a possible Nash equilibrium in stationary strategies, has a code indicating a player who would unilaterally deviate. For example the $\rightarrow$ in box, $a$, $r$, $u$ indicates that player 2 prefers $C_3(a, s, u)$ to $C_1(a, r, u)$. This is checked by observing that $\lambda_2(C_3) = 3 > \lambda_2(C_1) = 0$. The code "up" indicates a preference of $v$ over $u$ by player 3.

## 3. The "First Cycle" Game $F$

In this section we define a finite, $n$-person, perfect information game $F$, based on the same data as the DGA games (which is *not* a finite game) of the previous section. The game $F$ is played as before except that it stops as soon as a node is repeated. Each player $k$ then receives the payoff $\lambda_k^*(C)$, where $C$ is the first cycle to appear.

More formally, we define the extensive form (game tree) of $F$ as follows. Let $D \subset P$ be the set of all node-distinct paths and let $D_k = D \cap P_k$. Let $T \subset P$ denote the set of all paths $x = (1 = x_0, x_1, ..., x_r, ..., x_m)$ with $x^{m-1} \in D$ and $x_m = x_r$ where $m = l(x)$ and $r < m$. The paths $p \in T$ are the terminal nodes of the game tree of $F$, and the paths in $D$ are the non-terminal nodes. The payoff vector $f(x)$ assigned to an $x \in T$ is given by $\lambda^*(x_r, ..., x_m)$, the average local reward of the arcs of the first cycle $FC(x) = (x_r, x_{r+1}, ..., x_m)$ of $x$. Observe that the first cycle function $FC(x)$ is well defined for any $x$ in $\sim D \cup H$ (any history, or any path with a cycle), since every history contains a cycle.

A strategy for player $k$ in $F$ is a map $s_k\colon D_k \rightarrow N$ with $s_k(p)$ a successor node of $e(p) \in N_k$. A strategy profile $s = (s_1, ..., s_n)$ in $F$ determines a unique terminal "node" $p = \gamma(s) \in T$ in the same way as described in the last section, since the graph $G$ is finite and a node must eventually repeat. Using the same notation $f$ for payoff, define the payoff of a profile $s$ by $f(s) = f(\gamma(s))$. In other words the payoff vector of a profile is the average reward of the arcs of the first cycle it produces. It is important to note that the initial part of the game path not in the cycle is not relevant to the payoff.

Since $F$ is a finite perfect information game, the Zermelo–Kuhn procedure (see Owen [O, Sect. 1.4]) gives a perfect equilibrium strategy profile for $F$, which we will call $\hat{s}$. Let $\hat{p} = \gamma(\hat{s}) \in T$ be the terminal node corresponding to the equilibrium strategy profile $\hat{s}$ and let $\hat{C} = FC(\hat{p})$ be called the "equilibrium cycle." More generally, for any strategy profile $s$, let $C(s) = FC(\gamma(s))$ be the "terminal cycle" corresponding to $s$. Thus $\hat{C} = C(\hat{s})$ and the payoff is $f(s) = \lambda^*(C(s))$. In this notation, the equilibrium concept is described as follows:

THEOREM 1. *There is a finite recursive procedure yielding a strategy profile $\hat{s}$ for the "first cycle" game $F$, with the following property. Let $\hat{C} = FC(\gamma(\hat{s}))$. For any $k$ and strategy $s_k$ of player $k$, let $C(s_k) = FC(\gamma(\hat{s}_1, ..., \hat{s}_{k-1}, s_k, \hat{s}_{k+1}, ..., \hat{s}_n))$. Then $\lambda_k^*(C(s_k)) \leqslant \lambda_k^*(\hat{C})$.*

This simply says that if any player $k$ uniquely deviates from his equilibrium strategy, then the resulting terminal cycle is not better for him than the equilibrium cycle $\hat{C}$.

The game tree of the first cycle game $F$ corresponding to the complete graph game of Fig. 1 is shown in Fig. 2. The recursive algorithm for deter-

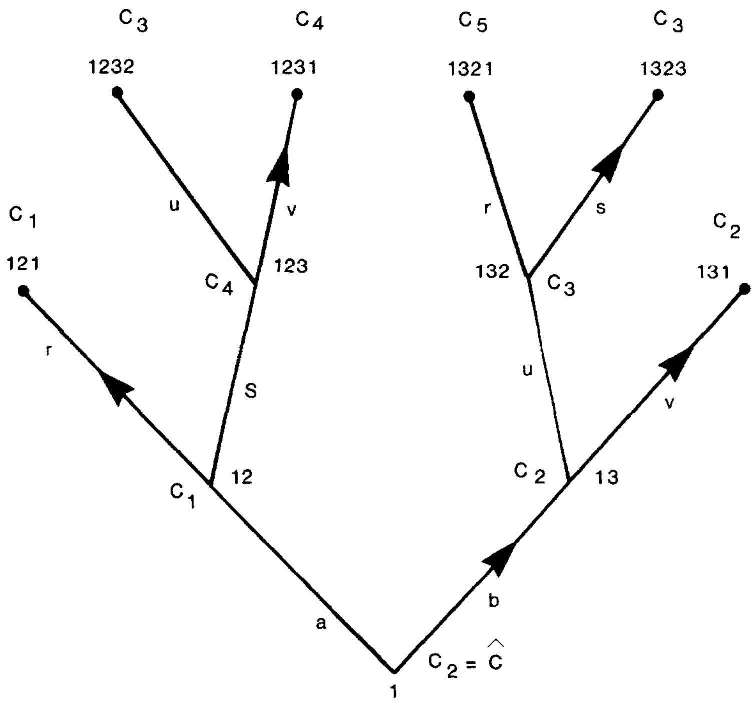


FIG. 2. The game tree $F$ for the DGA game of Fig. 1, and its recursive equilibrium $\hat{s}$, $\hat{C} = C_2$.

mining $\hat{s}$ is also indicated on the same figure. Initially outcomes (cycles $C_i$) are known only for the six terminal nodes. To determine the cycle at node 123, for example, player 3 must choose between $C_3$ (at node 1232) or $C_4$ (at node 1231). Since $\lambda_3^*(C_4)=0> -3/2 \;=\lambda^*(C_3)$, we have written $C_4$ at node 123, and drawn a "strategy arrow" on the arc $v$ from 123 to 1231. Eventually this finite process ends when $C_2$ is written at the starting node 1, at which point we learn that $C_2=\hat{C}$ is the equilibrium cycle. The strategy arrows determine the subgame perfect equilibrium strategy profile $\hat{s}$.

## 4. The Decycling Operator $d\colon P\to D$

In this section we define an operator $d$ which successively removes cycles from a path $x\in P$ until a node disjoint path $d(x)\in D$ is obtained. This operator will provide the link between an infinite DGA game $G$ and an associated "first cycle" game $F$.

Recall from the last section the first cycle operation $FC$ which assigns to any $x\in \sim D\cup H$ its first cycle $FC(x)$. For such an $x$ we define $\alpha(x)$ to be what remains after the first cycle of $x$ is removed. That is, if $x=(x_0, x_1, ..., x_r, ..., x_m, ...)$ where $x_m$ is the first repeated node, and $x_r=x_m$ with $r<m$, then $\alpha(x)=(x_0, x_1, ..., x_r,\; x_{m+1},\; x_{m+2}, ...)$ and $FC(x)=(x_r, ..., x_m)$.

For any path $x\in \sim D$ there is a least integer $q=\#(x)$, called the cycle number of $x$, such that $\alpha^q(x)\in D$ (has no more cycles). For $x\in \sim D$ define $d(x)=\alpha^{\#(x)}(x)$ and for $x\in D$ define $d(x)=x$. This defines $d$ for all $x\in P$ and we call $d(x)$ the decycled $x$. Observe that $d\colon P\to D$ and $d$ is the identity on $D$. For $x\in \sim D$ and $j=1, ..., \#(x)$ define cycles $C_j(x)$ inductively by $C_j(x)=FC(\alpha^{j-1}(x))$. For example, if $x=(1, 2, 3, 4, 3, 2, 4, 5)$ then $\#(x)=2$, $C_1(x)=(3, 4, 3)$, $C_2(x)=(2, 3, 2)$, and the decycled $x$ is $d(x)=(1, 2, 4, 5)$. For histories $h\in H$ we can define an infinite sequence of cycles $C_j(h)$, $j=1, 2, ...$, by $C_j(h)=FC(\alpha^{j-1}(h))$.

Observe that since any arc of a path $x\in \sim D$ is either in $\alpha(x)$ or in $FC(x)$, but not both, it is clear that we have the identity $\lambda(x)=\lambda(\alpha(x))+\lambda(FC(x))$. Applying this formula successively to the paths $\alpha^j(x)$, $j=1, ..., \#(x)-1$, we obtain the useful expansion

$$\lambda(x)=\lambda(d(x))+\sum_{j=1}^{\#(x)}\lambda(C_j(x)), \tag{4.1}$$

and

$$\lambda^*(x)=\frac{\lambda(d(x))}{l(x)}+\sum_{j=1}^{\#x}\frac{l(C_j(x))}{l(x)}\cdot\lambda^*(C_j(x)). \tag{4.2}$$

Similar but simpler reasoning gives us

$$l(x) = l(d(x)) + \sum_{j=1}^{\# x} l(C_j(x)). \tag{4.3}$$

## 5. Existence and Construction of Equilibrium Profiles

We now show how the constructions of the last two sections may be combined to construct equilibrium profiles for DGA games. Let $\hat{s} = (\hat{s}_1, ..., \hat{s}_n)$ be the equilibrium strategy for the "first cycle" game $F$, and let $d$: $P \to D$ be the decycling operator. Define a strategy profile $\hat{\sigma}$ for the DGA game by the equation

$$\hat{\sigma}_k(p) = \hat{s}_k(d(p)), \qquad k = 1, ..., n.$$

Observe that this is well defined because the operator $d$ is end-preserving, $e(d(p)) = e(p)$, and hence maps $P_k$ into $D_k$. The calculation of $\sigma_k$ involves two finite procedures: constructing the strategies $\hat{s}_k$ and computing $d(p)$, given $p$. The following two lemmas establish that $\hat{\sigma}$ is an equilibrium strategy profile, which will prove Theorem 2. Actually Lemma 2 proves that $\hat{\sigma}$ has equilibrium properties stronger than asserted in Theorem 2.

Definition. Let $W$ be any subset of the player set $\{1, ..., n\}$ and let $s_i$, $i \in W$, be strategies for player $i$ in the game $F$. A path $p \in D \cup T$ (i.e., a node of the game tree of $F$) is said to be *consistent* with $\{s_i\}_{i \in W}$ if for all $t < l(p)$ with $p_t \in N_i$ for some $i \in W$, we have $p_{t+1} = s_i(p^t)$.

Lemma 1. *Let $W$ be a subset of the players. Let $\sigma = (\sigma_1, ..., \sigma_n)$ be a strategy profile for the DGA game for which $\sigma_i = \hat{\sigma}_i$ for all $i \in W$. (Recall that $\hat{\sigma}_i(p) = \hat{s}_i(d(p))$.) Let $h = \infty(\sigma)$ be the history corresponding to $\sigma$. For $t = 0, 1, ...$ define $y(t) = (d(h^t), h_{t+1})$. Then for all $t \geqslant 0$, $y(t) \in D \cup T$ (a node of $F$) and $y(t)$ is consistent with the strategies $W^* = \{\hat{s}_i\}_{i \in W}$.*

*Proof.* To see that $y(t) \in D \cup T$ recall that the range of $d$ is $D$ so that $d(h^t) \in D$. Since $e(h^t) = e(d(h^t)) = h_t$ and $h_{t+1} \in \mathrm{Suc}(h_t)$ it follows that $y(t) \in P$. Hence $y(t)$ is either in $T$ or $D$ depending respectively on whether or not the node $h_{t+1}$ appears in the path $d(h^t)$.

To prove that for $t \geqslant 0$, $y(t)$ is consistent with $W^*$, we use induction on $t$. Observe that $y(0) = (h_0, h_1)$. If $h_0 = 1 \in P_i$ for some $i \in W$, then $h_1 = \sigma_i(h_0) = \hat{\sigma}_i(h_0) = \hat{s}_i(h_0)$, so $y(0)$ is consistent with $W^*$. Now assume that $y(m)$ is consistent with $W^*$. We consider separately the cases $y(m) \in D$

and $y(m) \in T$. If $y(m) \in D$ then $d(h^{m+1}) = (d(h^m), h_{m+1})$, and so $y(m+1) = (y(m), h_{m+2}) = (d(h^m), h_{m+1}, h_{m+2})$. If $h_{m+1} \in P_i$, $i \in W$, then $h_{m+2} = \hat{s}_i(y(m))$. So since $y(m)$ is consistent with $W^*$, by assumption, then so is $y(m+1)$. If $y(m) \in T$, then letting $b = l(y(m))$, we have $y(m)_r = y(m)_b = h_{m+1}$ for a unique $r < b$. In this case $d(h^{m+1}) = (y(m)_0, ..., y(m)^r$, and $y(m+1) = (y(m)^r, h_{m+2})$. Since $y(m)$ is consistent with $W^*$, so is the initial subpath $y(m)^r$. If $h_{m+1} \in P_i$ for some $i$ in $W$ then $h_{m+2} = \hat{s}_i(y(m)^r)$, so we also have that $y(m+1)$ is consistent with $W^*$.

LEMMA 2. *Let $\hat{h} = \infty(\hat{\sigma})$ be the history of the DGA game if every player $i$ adopts $\hat{\sigma}(p) = \hat{s}_i(d(p))$ and let $h(\sigma_k) = \infty(\hat{\sigma}_1, ..., \hat{\sigma}_{k-1}, \sigma_k, \hat{\sigma}_{k+1}, ..., \hat{\sigma}_n)$ be the history if all but player $k$ play as above and player $k$ plays $\sigma_k$. Then for every $j = 1, 2, ..., C_j(\hat{h}) = \hat{C}$, where $\hat{C}$ is the equilibrium cycle of $F$, and $\lambda_k^*(C_j(h(\sigma_k))) \leqslant \lambda_k^*(\hat{C})$.*

*Proof.* For any history $h$, let $t_1 < t_2 < \cdots < t_j <$ be the $t$'s such that $y(t) = (d(h^t), h_{t+1}) \in T$. Observe that for all $j$, $C_j(h) = FC(y(t_j))$. Take $h = \hat{h}$ and apply Lemma 1 with $W = \{1, ..., n\}$, the ful set of players. Hence the paths $y(t_j)$ are terminal nodes of $F$ which are consistent with $\{\hat{s}_i\}$ $i = 1, ..., n$. Hence $y(t_i) = \gamma(\hat{s})$ and so $C_j(\hat{h}) = FC(y(t_i)) = FC(\gamma(\hat{s})) = \hat{C}$.

Next, take $h = h(\sigma_k)$ and $W = \{1, ..., n\} - \{k\}$. It follows from Lemma 1 that $y(t_j)$ are terminal nodes of $F$ consistent with $\{\hat{s}_i\}_{i \neq k}$. Hence by Theorem 1, $\lambda_k^*(C_j(h(\sigma_k))) = \lambda_k^*(FC(y(t_j))) \leqslant \lambda_k^*(\hat{C})$.

THEOREM 2. *The recursively constructed profile $\hat{\sigma}$, given by $\hat{\sigma}_i(p) = \hat{s}_i(d(p))$, is an equilibrium strategy profile for the DGA game.*

*Proof.* Take $x = \hat{h}^t$ in (4.2), where $\hat{h} = \infty(\hat{\sigma})$, and noting that $C_j(x) = \hat{C}$ we obtain

$$\lambda^*(\hat{h}^t) = \frac{\lambda(d(\hat{h}^t))}{t} + \sum_{j=1}^{\#(\hat{h}^t)} \frac{l(\hat{C})}{t} \lambda^*(\hat{C}). \tag{5.1}$$

Since $\lambda$ is bounded on the finite set $D$ ($\ni d(\hat{h}^t)$) the first term goes to zero as $t \to \infty$. For the same reason, together with (4.3), the coefficient of $\lambda^*(\hat{C})$ in (5.1) goes to 1. Hence

$$\Pi(\hat{\sigma}) = \lim_{t \to \infty} \lambda^*(\hat{h}^t) = \lambda^*(\hat{C}). \tag{5.2}$$

Applying the same reasoning (and (4.2)) to $x = (h(\sigma_k))^t$, subtracting (5.1), and using the second part of Lemma 2, yields

$$\lambda_k^*(h(\sigma_k)^t) - \lambda_k^*(\hat{h}^t) \leqslant \frac{\lambda(d(h(\sigma_k)^t)) - \lambda(d(\hat{h}^t))}{t}. \tag{5.3}$$

Since the right side of (5.3) goes to zero,

$$\bar{\Pi}_k(\hat{\sigma}_1, ..., \hat{\sigma}_{k-1}, \sigma_k, \hat{\sigma}_{k+1}, \hat{\sigma}_N) \leqslant \overline{\lim}\, \lambda_k^*(h(\sigma_k)^t) \leqslant \Pi_k(\hat{\sigma}) = \hat{\lambda}_k^*(C). \tag{5.4}$$

Together, (5.2) and (5.4) establish the theorem.

If there are two players ($n=2$) and the local reward vector $\lambda$ has coordinates summing to zero, then we may interpret $\lambda_1(a)$ as the amount player 2 pays to player 1 when $a$ is traversed. We then define a "zero-sum DGA game" by taking 1 as maximizer and payoff $\bar{\Pi}_1$. (Alternatively we could have taken the lim inf evaluation $\underline{\Pi}_1$ as payoff.) It follows immediately from the previous theorem that this game has a minimax in pure strategies (as noted in the Introduction, the existence of stationary minimax strategies follows from known results).

COROLLARY 1. *If $\lambda_1 + \lambda_2$ is identically zero on $A$ then there exist strategies $\bar{\sigma}_1 \in S_1$ and $\bar{\sigma}_2 \in S_2$ such that*

$$\bar{\Pi}_1(\sigma_1, \bar{\sigma}_2) \leqslant \Pi_1(\bar{\sigma}_1, \bar{\sigma}_2) \leqslant \bar{\Pi}_1(\bar{\sigma}_1, \sigma_2) \qquad \textit{for all} \quad \sigma_1 \in S_1, \textit{ and } \sigma_2 \in S_2.$$

*Proof.* Let $\bar{\sigma} = (\bar{\sigma}_1, \bar{\sigma}_2)$ be any equilibrium profile, as guaranteed by Theorem 2. The first inequality is the definition of equilibrium, with $k=1$. The hypothesis on $\lambda$ gives $\bar{\Pi}_2 = -\underline{\Pi}_1$ and $\Pi_2 = -\Pi_1$ (if the limits exist). Hence, taking $k=2$ in the definition of equilibrium gives

$$\begin{aligned} \bar{\Pi}_1(\bar{\sigma}_1, \sigma_2) &\geqslant \underline{\Pi}_1(\bar{\sigma}_1, \sigma_2) \\ &= -\bar{\Pi}_2(\bar{\sigma}_1, \sigma_2) \\ &\geqslant -\bar{\Pi}_2(\bar{\sigma}_1, \bar{\sigma}_2) \\ &= \bar{\Pi}_1(\bar{\sigma}_1, \bar{\sigma}_2). \end{aligned}$$

## 6. EXISTENCE OF AUTOMATED STRATEGY EQUILIBRIA

The results of Sections 2 and 5 show that while there are always DGA equilibria using strategies depending on the entire past history, there may well be no equilibrium in very simple, say stationary, strategies. In this section we demonstrate that there are always equilibrium profiles employing strategies implementable by finite state automata. Such strategies have recently been the focus of much attention in game theory, particularly for repeated games. (See [R].)

We define an $m$-automated strategy for player $k$ in a DGA game as a 3-tuple $(Z_k, \alpha_k, \beta_k)$ where $Z_k$ is a set (of internal states) with no more than $m$ elements, $\alpha_k\colon Z_k \times N \to Z_k$ and $\beta_k\colon Z_k \times N \to N$ are maps. If $z_0 \in Z_k$ is the initial internal state, and $x$ is a history of the game consistent with this strategy then we define $z_t = \alpha(z_{t-1}, x_t)$ for $t = 1, 2, \ldots$ and require that $\beta(z_{t-1}, x_t) \in \mathrm{Suc}(x_t)$ whenever $e(x^t) \in N_k$.

The interpretation of the above definition is straightforward. As the game proceeds the new node $x_t$ is input into player $k$'s automaton, which then changes its internal state accordingly. If $x_t$ is not a player $k$ node, the automaton's output node $w_{t+1} = \beta_k(z_{t-1}, x_t)$ is ignored. However, if $x_t \in N_k$ then the referee takes the output $w_{t+1} = x_{t-1}$ as the next node. As an example, a stationary strategy is a 1-automated strategy, with $Z_k = \{z_0\}$, $\alpha_k$ the constant $z_0$, and $\beta_k(z_0, i)$ is the unique response to the node $i$. Thus the example in Section 2 demonstrates that, in general, there need not be an equilibrium in 1-automated strategies. However, we will prove the following:

THEOREM 3. *In any DGA game with node set $N$ there is an equilibrium strategy profile in $(\#N-1)!$-automated strategies $(Z_k, \alpha_k, \beta_k)$, $k = 1, \ldots, n$.*

*Proof.* We will show that the strategies $\hat{\sigma}_k(p) = \hat{s}_k(d(p))$, $k = 1, \ldots, n$, can in fact be implemented by automated strategies. Since the $\hat{\sigma}_k$ form an equilibrium profile (Theorem 2), this will prove the theorem.

The recursive computation of the $\hat{\sigma}_k$ is a simple consequence of the identity

$$d(x^t) = d(d(x^{t-1}), x_t) \qquad \text{for} \quad x \in P, \quad l(x) \geqslant t. \tag{6.1}$$

Let $Z_k = D$, with $z_0 = 1$ (the initial node) and observe that $\#(D) \leqslant (\#N-1)!$ with equality only for the complete graph. For $x \in D$ and $i \in \mathrm{Suc}(e(p))$ define $\alpha_k(p, i) = d(p, i)$ and $\beta_k(p, i) = s_k(d(p, i))$. If $i \notin \mathrm{Suc}(e(p))$ then $\alpha_k(p, i)$ and $\beta_k(p, i)$ may be defined arbitrarily. Let $z_0 = 1 = x_0$ and defined $x_1, z_1, x_2, z_2, \ldots$ recursively as follows. If $x_t \in N_j$, $j \neq k$, then player $j$ chooses $x_{t+1} = \sigma_j(x^t)$ where $\sigma_j \in S_j$. If $x_t \in N_k$ then $x_{t+1} = \beta_k(z_t, x_t)$. Similarly $z_{t+1} = \alpha_k(z_t, x_{t+1})$. It follows inductively from these definitions and (6.1) that $z_t = d(x^t)$ for all $t$. Consequently when $x_t \in N_k$ we have that $x_{t+1} = \beta_k(z_{t-1}, x_t) = \hat{s}_k(d(d(x^{t-1})), x^t) = \hat{s}_k(d(x^t)) = \hat{\sigma}_k(x^t)$. Hence $(Z_k, \alpha_k, \beta_k)$ implements $\hat{\sigma}_k$, as required, and the theorem is proved.

## 7. AN APPLICATION TO DUOPOLY PRICING

We now give an example of a simple economic problem which may be viewed as a DGA game on a complete bipartite graph. Consider a duopoly model where two firms I and II alternately set prices which they cannot then change for two time periods. In any period when I's price is indexed by $i$ (chosen from the finite set $\{1, ..., m\}$) and II's price is $j \in \{1, ..., n\}$ the respective profits for I and II are $A(i, j)$ and $B(i, j)$, where $A$ and $B$ are known $m \times n$ matrices. The game begins in period 1 where II chooses his price $j_1$ knowing I's externally fixed initial prices $1 = i_1$. In succeeding periods they alternately choose prices $i_2, j_2, i_3, ..., i_t, j_t, ...$ with knowledge of all preceding prices. The average profit for II for the price sequence up to $j_t, i_{t+1}$ is given by

$$\frac{1}{2t} \sum_{n=1}^{t} [B(i_t, j_t) + B(i_{t+1}, j_t)]$$

with a similar formula with $A$ replacing $B$ for I's average profit. We assume that each player wishes to maximize his long term average profit. Similar models have appeared in the economics literature for players maximizing the discounted sum of their single period profits. (See [MT, W2].)

The above problem may be modeled as a DGA game on a complete bipartite graph $G$ whose nodes are partitioned into two sets $N_{\mathrm{I}} = \{a_1, ..., a_n\}$ and $N_{\mathrm{II}} = \{b_1, ..., b_m\}$ and with arcs

$$A = \{(a_j, b_i), (b_i, a_j): i = 1, ..., m; j = 1, ..., n\}.$$

The node $a_j$ is interpreted as the situation where I plays after II has chosen price $j$. The distinguished starting node 1 is $a_1$ and the local reward vector $\lambda: A \to \mathbb{R}^2$ is given by $\lambda((a_j, b_i)) = (A(i, j), B(i, j)) = \lambda((b_i, a_j))$. The resulting DGA game may be recursively solved using the techniques of Sections 3–5. We leave the explanation and interpretation of the equilibrium cycle $\hat{C}$ to the economists.

## 8. Recapture Games and Surveillance Games

In this section we define two new classes of games ("recapture" and "surveillance") which can be modeled as DGA games. These games are related to the "discrete differential games" introduced by Isaacs [I, Chap. 3], including the "hamstrung squad car game" and the "homicidal chauffeur game." All the games considered in this section are two person zero-sum. In Isaacs' games two players, called the pursuer and the evader, alternately move their positions along a given graph according to their respective "rules of motion." The payoff to the maximizing evader is the number of moves $T$ until their positions coincide ($T$ is called the capture time). Isaacs also considered "games of kind" in which the evader wins if he avoids capture forever and the pursuer wins on capture.

Our two classes of games have the same dynamics as Isaac's games except that play goes on indefinitely, even after capture. The "recapture game" has a payoff to the maximizing pursuer of the long term fraction of the time that the two players share the same position (node). Equivalently the payoff may be taken as the average time taken for the pursuer to recapture the evader, with the pursuer as minimizer, since this is the reciprocal of the previous payoff. We also define a "surveillance game" where the payoff to the maximizing evader is the long term average distance between the players. This game is a perfect information analog of the game of Cohen, Chung, and Graham [CCG] where the payoff is the graph distance between nodes picked simultaneously by two players from a given graph.

We now formally describe the common dynamics needed to define recapture and surveillance games as DG games. Let $N$ denote the common set of nodes to be occupied by the players, and let $A_k \subset N \times N$, $k = 1, 2, ...,$ denote the arc sets corresponding to the rules of motion for the evader $k = 1$ and the pursuer $k = 2$. The DG dynamics are then given by the data $(\bar{N}, \bar{N}_1, \bar{N}_2, \bar{A})$, where $\bar{N} = N \times N \times \{1, 2\}$, $\bar{N}_k = N \times N \times \{k\}$, and

$$\bar{A} = \{((i, j, 1), (i', j, 2)) \colon (i, i') \in A_1\} \cup \{((i, j, 2), (i, j', 1)) \colon (j, j') \in A_2\}.$$

The node $(i, j, k)$ describes the situation where the evader occupies node $i$, the pursuer is at node $j$, and it is player $k$'s turn to move.

To make these DG dynamics into DGA games of recapture or surveillance we must define appropriate local reward vectors $\lambda$. Actually it will be easier to define local rewards on nodes rather than arcs. The former method is a special case of the latter, since we may induce the reward from a node onto all arcs leaving that node. Since both games are zero sum we need only specify the local reward to the maximizer. For the recapture game we define $\lambda \colon \bar{N} \to \mathbb{R}$ by $\lambda(i, j, k) = \sigma_{ij}$ (1 if $i = j$; otherwise 0). For any history $h$, $\Pi(h)$ is the fraction of time the players occupy the same node, and $1/\Pi(h)$ is the average recapture time (if these limits exist—they do exist at equilibrium). To define the payoff for the surveillance game, fix any metric $\rho$ on $N$, for example the graph distance between nodes for some undirected graph. Then define $\lambda \colon \bar{N} \to \mathbb{R}$ by $\lambda(i, j, k) = \rho(i, j)$. The payoff $\Pi(h)$ is now the long term average distance between the two players.

To make these games more specific we consider the following dynamics which we call the "cyclic dynamics $m$, $M_1, M_2$": Let $C(2m)$ be the cyclic graph consisting of nodes $N = \{0, 1, ..., 2m - 1\}$ arranged in the usual order around a circle, with $\rho$ the graph distance between nodes. For $M_1$ and $M_2$ arbitrary subsets of $\{0, 1, ..., m - 1\}$, define $A_k = \{(i, j) \in N \times N \colon \rho(i, j) \in M_k\}$, $k = 1, 2$, so that the players' rules of motion are simply allowable step-lengths.

Now consider the DG cyclic dynamics with $m \geqslant 4$, $M_1 = \{0, 1\}$, and $M_2 = \{2, 3\}$. This means that the evader can stay still or move one node in either direction, while the pursuer must move in steps of length 2 or 3. We claim that the recapture game has value 0 because the evader can successfully avoid recapture forever. To describe this evasion strategy we must first observe how strategies in cyclic dynamics can be denoted. Since the essence of the situation is given simply by the distance $\rho$ between the players ($\rho$ is a "reduced state variable" in Isaac's terminology) a strategy for player $k$ is a rule $\rho \to \rho'$ ($\rho'$ is distance *after* player $k$ moves) with $\rho' = |\rho \pm \varepsilon|$, some $\varepsilon \in M_k$. The successful evasion strategy is now defined by $0 \to 1$ ($\varepsilon = 1$), $1 \to 1$ ($\varepsilon = 0$), $2 \to 1$ ($\varepsilon = 1$), and $\rho' = \rho + 1$ for $\rho \geqslant 3$. In the surveillance game the maximizing hider guarantees an average distance of at least 5/4 by the strategy $\rho' = 0$ if $\rho \leqslant 1$ and $\rho' = \rho + 1$ if $\rho \geqslant 2$. Similarly the pursuer guarantees $\leqslant 5/4$ by playing $\rho' = 0$ if $\rho = 2$, $\rho' = 1$ if $\rho = 1$, $\rho' = 2$ if $\rho = 0$, and $\rho' = \rho - 3$ if $\rho \geqslant 3$. The optimal cycle of distance is then 0, 0, 2, 3, 0, 0, 2, 3, ... and the value is 5/4.

Now consider the cyclic dynamics where $m \geqslant 4$, $M_1 = \{0, 1\}$ (as before), and $M_2 = \{2, 3, 4\}$ (the step-length 4 has been added). The analysis of the surveillance game is unchanged, but in the recapture game the pursuer can now guarantee capture, so the value is positive. In particular the value is 1/4, and the mean recapture time is 4. The optimal pursuer strategy (in distance notation) is $0 \to 3$, $1 \to 3$, $2 \to 0$, $3 \to 0$, and $\rho' = \rho - 4$ for $\rho \geqslant 4$. An optimal evader strategy is to always move away (1 step) from the pursuer. The distance sequence following capture is given by 0, 1, 3, 4, 0, 1, 3, 4, ..., for which $\Pi = 1/4$. Again we leave the easy proofs, that these strategies are optimal, to the reader.

We conclude this section with two discussions about generalizing recapture and surveillance games with regard to dynamics (rules of motion) and payoffs. Note that our model does not at present allow dynamics that depend on history. The "hamstrung squad car game" [I, p. 57; BB] is played on a rectangular lattice representing city streets, where the evader can move to any of the four adjacent lattice points and the pursuer can move either two "ahead" or two "to the right" because he is a policeman and must follow traffic laws prohibiting left or $U$-turns. Capture takes place when the players occupy adjacent lattice points. In order for us to have dynamics such as "ahead" or "right" we must code the previous as well as current node of the pursuer in $\bar{N}$. This can be done by defining $\bar{N} = N \times N \times N \times \{1, 2\}$ where $(i, j, l, k)$ has the additional content that the pursuer is current at node $j$ but previously was at node $l$. Our second generalization involves payoffs. So far in this section we have for simplicity considered only two person zero-sum games. We could just as well consider a two person game where the evader has the recapture payoff and the pursuer has the surveillance payoff, for example. Moreover we can consider three person games where the local payoff vector when 1 is at $i$, 2 at $j$, 3 at $r$ is given by

$$\lambda(i, j, r) = (\rho(i, r) - \rho(i, j), \rho(j, i) - \rho(j, r), \rho(r, j) - \rho(r, i)).$$

Equilibrium strategies for this highly unstable "triangle game" are guaranteed by Theorem 2.

## 9. Cycles in Imperfect Information Games

We conclude this paper with some remarks on allowing cycles in extensive form games of imperfect information. Here we use the traditional definition of pure and behavioral strategies as deterministic or probabilistic maps defined on a player's information sets. If we consider the infinite tree form of the game it is clear that every play of the game will pass through some information set infinitely often. Hence the traditional assumption that an information set may be entered at most once in any play is violated. However, the way to deal with such games (at least for finite trees) was shown by Isbell [Is]: Players must randomize over a finite number of behavioral strategies, and Nash equilibria in such strategies always exist. (The author discovered this fact much later [A] in the course of these investigations.) We conjecture that imperfect information games with cycles have Nash equilibria in general (not finitely supported) distributions over behavioral strategies. To avoid possible confusion it should be noted that stationary strategies in DGA games correspond to what we are calling pure strategies here. In this context the example of Section 2 shows that when cycles are allowed, extensive form games with singleton information sets need not have pure strategy Nash equilibria. There is room for much further research in this area.

## References

[A] S. Alpern, Games with repeated decisions. *SIAM J. Control Optim.* **26**, No. 2 (1988), 468–477.

[B1] C. Berge, Topological games with perfect information, *Contrib. Theory Games* **3** (1957), 165–178.

[B2] C. Berge, "The Theory of Graphs," Methuen, London, 1962.

[BB] V. Baston and F. Bostock, Hammstrung squad car games, *Internat. J. Game Theory* **16**, No. 4 (1987), 253–261.

[$BB_1$] V. Baston and F. Bostock, "On Washburn's Deterministic Graphical Games," Proceedings of the Fourth International Symposium on Differential Games and Applications, Helsinki, 1990.

[$BB_2$] V. Baston and F. Bostock, "Infinite Deterministic Graphical Games," preprint (1991).

[BK] T. Bewley and E. Kohlberg, On stochastic games with stationary optimal strategies, *Math. Oper. Res.* **3**, No. 2 (1978), 104–125.

[CCG] F. R. K. Chung, J. E. Cohen, and R. L. Graham, Pursuit-evasion games on graphs, *J. Graph Theory* **12**, No. 2 (1988).

[E] H. Everett, Recursive games, *in* "Contributions to the Theory of Games," Vol. III, p. 47–78, Annals of Mathematical Studies, Vol. 39, Princeton Univ. Press, Princeton, NJ, 1957.

[EM] A. Ehrenfeucht and J. Mycielski, Positional strategies for mean payoff games, *Internat. J. Game Theory* **8** (1979), 109–113.

[G] D. Gillette, Stochastic games with zero stop probabilities, *in* "Contributions to the Theory of Games," Vol. 3, Annals of Mathematical Studies, Vol. 39, Princeton Univ. Press, Princeton, NJ, 1957.

[I] R. Isaacs, "Differential Games," Wiley, New York, 1965.

[Is] J. Isbell, Finitary games, *in* "Annals of Mathematical Studies," Vol. 39, pp. 79–95, Princeton Univ. Press, Princeton, NJ, 1957.

[LL] T. Liggett and S. Lippman, Stochastic games with perfect information and time average payoff, *SIAM Rev.* **2** (1969), 604–607.

[MT] E. Maskin and J. Tirole, A theory of dyamic oligopoly. I, II. Price competition, kinked demand curves, and Edgeworth cycles, *Econometrica* **56**, No. 3 (1988), 549–599.

[O] G. Owen, "Game Theory," Academic Press, New York, 1982.

[R] A. Rubinstein, Finite automata play the repeated prisoner's dilemma, *J. Econ. Theory* **39** (1986), 83–96.

[W1] A. Washburn, Deterministic graphical games, preprint.

[W2] A. Washburn, A computation method for matrix differential games, *Internat. J. Game Theory* **6**, No. 1 (1977), 1–10.